On the State Complexity of the Reverse of $\mathcal{R}$- and $\mathcal{J}$-trivial Regular Languages

Galina Jirásková [1,⋆] and Tomáš Masopust [2,⋆⋆]

[1] Mathematical Institute, Slovak Academy of Sciences
Grešákova 6, 040 01 Košice, Slovak Republic
jiraskov@saske.sk

[2] Institute of Mathematics, Academy of Sciences of the Czech Republic
Žižkova 22, 616 62 Brno, Czech Republic
masopust@math.cas.cz

Abstract. The tight upper bound on the state complexity of the reverse of $\mathcal{R}$-trivial and $\mathcal{J}$-trivial regular languages is 2^{n-1} for languages of the state complexity n. The witness for $\mathcal{R}$-trivial regular languages is over a ternary alphabet while for $\mathcal{J}$-trivial regular languages over an $(n-1)$-element alphabet. In this paper, we prove that the size of the alphabets cannot be improved, that is, the bound can be met neither by a binary $\mathcal{R}$-trivial regular language nor by a $\mathcal{J}$-trivial regular language over an $(n-2)$-element alphabet. We also present tight bounds for binary $\mathcal{R}$-trivial and $(n-2)$-element $\mathcal{J}$-trivial regular languages. Tight bounds for $(n-k)$-element $\mathcal{J}$-trivial regular languages, for $2 \le k \le n-3$, are open.

1 Introduction

Regular languages of simple forms play an important role in mathematics and computer science. The reader is referred to, e.g., [1,7,15] for a few applications of $\mathcal{J}$-trivial or piecewise testable languages.

The state complexity of a regular language is the number of states of a minimal deterministic finite automaton (DFA) accepting the language. Equivalently, the state complexity coincides with the quotient complexity [4], which is the number of distinct (left) quotients of a regular language.

The reverse of a machine or of a language is one of the classical operations in automata and formal language theory. However, in comparison with, e.g., boolean operations, the state complexity of the reverse of regular languages is exponential in the worst case, and there exist binary witness languages of the state complexity n with the reverse of the state complexity 2^n, see [14,20]. This even holds true for union-free regular languages defined by regular expressions without the union operation [9].

In this paper, we consider languages defined by Green's equivalence relations, namely $\mathcal{R}$-trivial and $\mathcal{J}$-trivial regular languages. Let M be a monoid and s and

⋆ Research supported by VEGA grant 2/0183/11 and by grant APVV-0035-10.
⋆⋆ Research supported by GAČR grant P202/11/P028 and by RVO: 67985840.

t be two elements of M. Green's relations $\mathcal{L}$, $\mathcal{R}$, $\mathcal{J}$, and $\mathcal{H}$ on M are defined as follows: $(s,t) \in \mathcal{L}$ if and only if $M \cdot s = M \cdot t$, $(s,t) \in \mathcal{R}$ if and only if $s \cdot M = t \cdot M$, $(s,t) \in \mathcal{J}$ if and only if $M \cdot s \cdot M = M \cdot t \cdot M$, and $\mathcal{H} = \mathcal{L} \cap \mathcal{R}$. For $\rho \in \{\mathcal{L}, \mathcal{R}, \mathcal{J}, \mathcal{H}\}$, M is ρ-trivial if $(s,t) \in \rho$ implies $s = t$, for all s,t in M. A language is *ρ-trivial* if its syntactic monoid is ρ-trivial. Note that $\mathcal{H}$-trivial regular languages coincide with *star-free* languages [13, Chapter 11] and $\mathcal{L}$-trivial, $\mathcal{R}$-trivial and $\mathcal{J}$-trivial regular languages are all star-free languages. Moreover, $\mathcal{J}$-trivial regular languages are both $\mathcal{L}$-trivial and $\mathcal{R}$-trivial.

From the computer science viewpoint, a regular language is $\mathcal{R}$-trivial if and only if it is a finite union of languages of the form $\Sigma_1^* a_1 \Sigma_2^* a_2 \Sigma_3^* \cdots \Sigma_k^* a_k \Sigma^*$, where $k \geq 0$, $a_i \in \Sigma$, and $\Sigma_i \subseteq \Sigma \setminus \{a_i\}$, or if and only if it is accepted by a partially ordered minimal DFA [3]. Similarly, a regular language is $\mathcal{J}$-trivial (also called *piecewise testable*) if and only if it is a finite boolean combination of languages of the form $\Sigma^* a_1 \Sigma^* a_2 \Sigma^* \ldots \Sigma^* a_k \Sigma^*$, where $k \geq 0$ and $a_i \in \Sigma$, or if and only if the minimal DFAs for both the language and the reverse of the language are partially ordered [16,17]. Other automata representations of these languages can be found, e.g., in [8] and the literature therein.

In 1985, Stern [18] suggested a polynomial-time algorithm of order $O(n^5)$ in the number of states and transitions of the minimal DFA to decide whether a regular language is $\mathcal{J}$-trivial. Recently, Trahtman [19] improved this result to a quadratic-time algorithm. Polák and Klíma [12] have further investigated the properties of piecewise testable languages and introduced so-called biautomata that can be used to verify piecewise testability of a regular language, but which can be of exponential size in the worst case as shown in [11].

In [10], we have shown that the upper bound on the state complexity of the reverse of $\mathcal{R}$-trivial and $\mathcal{J}$-trivial regular languages is 2^{n-1} for languages of the state complexity n and that this bound can be met by a ternary $\mathcal{R}$-trivial regular language. Furthermore, we have conjectured that an $(n-1)$-element alphabet is sufficient for $\mathcal{J}$-trivial regular languages of the state complexity n to meet the upper bound. A proof can be found in [5].

In this paper, we prove the optimality of the alphabets. Namely, we prove that the bound on the state complexity of the reverse can be met neither by a binary $\mathcal{R}$-trivial regular language nor by a $\mathcal{J}$-trivial regular language over an $(n-2)$-element alphabet. We also show tight upper bounds for binary $\mathcal{R}$-trivial regular languages. As a result, we give a characterization of tight upper bounds for $\mathcal{R}$-trivial regular languages depending on the state complexity of the language and the size of the alphabet. Finally, we prove a tight upper bound for $(n-2)$-element $\mathcal{J}$-trivial regular languages and a few tight bounds for binary $\mathcal{J}$-trivial regular languages. The case of $(n-k)$-element $\mathcal{J}$-trivial regular languages, for $2 \leq k \leq n-3$, is left open.

2 Preliminaries and Definitions

We assume that the reader is familiar with automata and formal language theory [13]. The cardinality of a set A is denoted by $|A|$, and the powerset of A is

denoted by 2^A. An alphabet is a finite nonempty set. The free monoid generated by an alphabet Σ is denoted by Σ^*. A string over Σ is any element of Σ^*. The empty string is denoted by ε.

A *nondeterministic finite automaton* (NFA) is a 5-tuple $M = (Q, \Sigma, \delta, Q_0, F)$, where Q is the finite nonempty set of states, Σ is the input alphabet, $Q_0 \subseteq Q$ is the set of initial states, $F \subseteq Q$ is the set of accepting states, and $\delta : Q \times \Sigma \to 2^Q$ is the transition function that can be extended to the domain $2^Q \times \Sigma^*$. The language *accepted* by M is the set $L(M) = \{w \in \Sigma^* \mid \delta(Q_0, w) \cap F \neq \emptyset\}$. The NFA M is *deterministic* (DFA) if $|Q_0| = 1$ and $|\delta(q, a)| = 1$ for every q in Q and a in Σ. In this case we identify singleton sets with their elements and simply write q instead of $\{q\}$. Moreover, we consider the transition function δ to be a total mapping from $Q \times \Sigma$ to Q that can be extended to the domain $Q \times \Sigma^*$. A non-accepting state $d \in Q$ such that $\delta(d, a) = d$, for all a in Σ, is called a *dead* state. Two states of a DFA are *distinguishable* if there exists a string w that is accepted from one of them and rejected from the other; otherwise they are *equivalent*. A DFA is *minimal* if all its states are reachable and pairwise distinguishable.

A DFA $M = (Q, \Sigma, \delta, Q_0, F)$ is *partially ordered* if the reachability relation $\preceq$ on the set of states, defined by $p \preceq q$ if there exists a string w in Σ^* such that $\delta(q, w) = p$, is a partial order.

The *subset automaton* of an NFA $M = (Q, \Sigma, \delta, Q_0, F)$ is the DFA $M' = (2^Q, \Sigma, \delta', Q_0, F')$ constructed by the standard subset construction.

The *reverse* w^R *of a string* w is defined by $\varepsilon^R = \varepsilon$ and $(va)^R = av^R$, for v in Σ^* and a in Σ. The *reverse of a language* L is the language $L^R = \{w^R \mid w \in L\}$. The *reverse of a DFA* M is the NFA M^R obtained from M by reversing all transitions and swapping the role of initial and accepting states.

The following result says that there are no equivalent states in the subset automaton of the reverse of a minimal DFA. We use this fact in the paper when proving the tightness of upper bounds. By this fact, it is sufficient to show that the corresponding number of states is reachable in the subset automaton since the distinguishability always holds.

Fact 1 ([2]) *All states of the subset automaton corresponding to the reverse of a minimal DFA are pairwise distinguishable.* □

See Appendix, p. 13, for a proof.

In this paper we implicitly use the characterizations that a regular language is $\mathcal{R}$-trivial if and only if it is accepted by a minimal partially ordered DFA and that it is $\mathcal{J}$-trivial if and only if both the language and its reverse are accepted by minimal partially ordered DFAs.

Note that this immediately implies that $\mathcal{J}$-trivial regular languages are closed under reverse. However, $\mathcal{R}$-trivial regular languages are not closed under reverse since the $\mathcal{R}$-trivial regular language of Fig. 1 in [10] is not $\mathcal{J}$-trivial, hence the minimal DFA for the reverse is not partially ordered.

The *state complexity* of a regular language L, denoted by $\mathrm{sc}(L)$, is the number of states in the minimal DFA recognizing the language L.

The following lemma shows that in some cases we do not need to distinguish between DFAs with and without dead state. In particular, we can get the result

for DFAs without a dead state immediately from an analogous result for DFAs with a dead state or vice versa.[3]

Lemma 1. *Let L be a regular language. Then* $\mathrm{sc}(L) = \mathrm{sc}(L^c)$*, where* L^c *denotes the complement of* L*. In particular, we have* $\mathrm{sc}(L^R) = \mathrm{sc}((L^c)^R)$.

Proof. Let M be a minimal DFA accepting L. Then M^c constructed from M by swapping accepting and non-accepting states is a minimal DFA accepting L^c. The second part now follows by the observation that $(L^R)^c = (L^c)^R$. □

This lemma says that if a DFA M has a dead state and reaches the upper bound, then the same result for DFAs without a dead state can be derived from the complement of M if it does not have a dead state. However, note that Table 1 demonstrates that there are cases where this technique fails because both the DFA and its complement have a dead state.

Immediate consequences of this lemma combined with the known results are formulated below.

Corollary 1. *(i) There exist ternary* $\mathcal{R}$*-trivial regular languages* L_1 *and* L_2 *whose automata representation has and does not have a dead state, respectively, with* $\mathrm{sc}(L_1) = \mathrm{sc}(L_2) = n$ *and* $\mathrm{sc}(L_1^R) = \mathrm{sc}(L_2^R) = 2^{n-1}$*. (ii) There exist* $\mathcal{J}$*-trivial regular languages* L_1 *and* L_2 *over an alphabet* Σ *with* $|\Sigma| \geq n-1$ *whose automata representation has and does not have a dead state, respectively, with* $\mathrm{sc}(L_1) = \mathrm{sc}(L_2) = n$ *and* $\mathrm{sc}(L_1^R) = \mathrm{sc}(L_2^R) = 2^{n-1}$.

See Appendix, p. 13, for the figures.

Proof. Using Lemma 1, (i) follows from [10, Lemma 3, p. 232] since the automaton used there has a dead state and its complement does not, while (ii) follows from the automaton used in [5, v2, Theorem 5, p. 15]. □

3 $\mathcal{R}$-trivial regular languages

Recall that the state complexity of the reverse of $\mathcal{R}$-trivial regular languages with the state complexity n is 2^{n-1} and there exists a ternary witness language meeting the upper bound. We now prove that the ternary alphabet is optimal, that is, the upper bound cannot be met by any binary $\mathcal{R}$-trivial regular language.

Lemma 2. *Let L be a binary* $\mathcal{R}$*-trivial regular language with* $\mathrm{sc}(L) = n$*, where* $n \geq 2$*. Then* $\mathrm{sc}(L^R) \leq 2^{n-2} + n - 1$.

Proof. Let $M = (\{1, \dots, n\}, \{a, b\}, \delta, 1, F)$ be a partially ordered DFA with n states such that $i \preceq j$ implies $i \leq j$. Let M' be the subset automaton of the NFA M^R. We show that M' has at most $n-1$ reachable sets that contain $n-1$. By Lemma 1, we can assume that state n is accepting, otherwise we take the complement. Then there are two cases: (i) both symbols a, b go from state $n-1$ to state n, or (ii) without loss of generality, the transition under b goes from $n-1$ to n and there is a self-loop under a in state $n-1$.

[3] We are grateful to an anonymous referee for pointing out this observation.

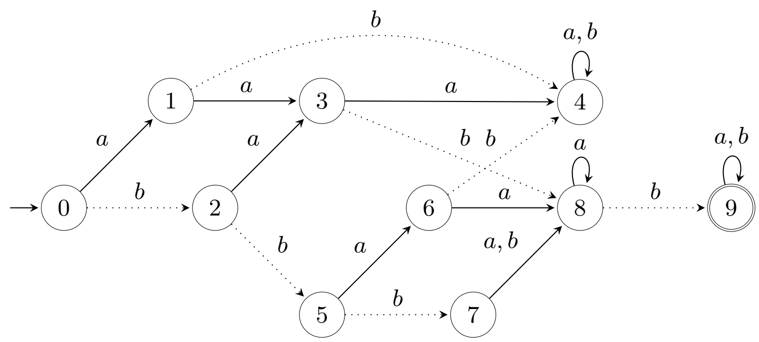

Fig. 1. There are three trees, namely $T_4 = \{0, 1, 2, 3, 4\}$, $T_8 = \{5, 6, 7, 8\}$, and $T_9 = \{9\}$; b-transitions are dotted.

In the first case, no sets without state $n-1$ are reachable in M', except for F. Hence the upper bound is at most $2^{n-2}+1$.

In the second case, all sets not containing state $n-1$ must be reachable by strings in a^*. We now prove that at most $n-1$ such sets are reachable in M'. It is sufficient to prove that $F \cdot a^{n-1} = F \cdot a^{n-2}$, where $\cdot$ is the transition function of the subset automaton M'.

The subautomaton of M, defined by restricting to the alphabet $\{a\}$, is a disjoint union of trees T_q where $\delta(q,a) = q$ and T_q consists of all states that can reach q by a string in a^*; see Fig. 1 for illustration. Let k be the depth of T_q, and let $F' = F \cap T_q$. If $q \in F'$, then $F' \cdot a^k = T_q$. If $k \notin F'$, then $F' \cdot a^k = \emptyset$. In both cases $F' \cdot a^k = F' \cdot a^{k+1}$. Now $F \cdot a^m$ is a disjoint union of such $F' \cdot a^m$. By the assumptions, all trees are of depth at most $n-2$; recall that there is no transition under a from $n-1$ to n. Hence $F \cdot a^{n-1} = F \cdot a^{n-2}$ follows. □

The following lemma proves the lower bound 2^{n-2} on the state complexity of the reverse of binary $\mathcal{R}$-trivial regular languages.

Lemma 3. *For every $n \geq 3$, there exists a binary $\mathcal{R}$-trivial regular language with* $\mathrm{sc}(L) = n$ *such that* $\mathrm{sc}(L^R) \geq 2^{n-2}$.

Proof. Consider the language L accepted by the partially ordered binary n-state DFA M shown in Fig. 2. We now show that each subset of $\{0, 1, \ldots, n-2\}$ is reachable in the subset automaton of the NFA M^R.

The proof is by induction on the size of subsets. The subset $\{0\}$ is the initial state of the subset automaton. Each subset $\{0, i_1, i_2, \ldots, i_k\}$ of size $k+1$, where $1 \leq i_1 < i_2 < \cdots < i_k \leq n-1$, is reached from the subset $\{0, i_2 - i_1, \ldots, i_k - i_1\}$ of size k by the string ab^{i_1-1}. □

Using a computer program we have computed a few tight bounds summarized in Table 1. Note that the upper bound $2^{n-2} + (n-1)$ is met by a DFA for L if $n \leq 6$, but not if $n = 7$. In addition, more than 2^{n-2} states are reachable if

$n =$ sc(L)	Worst-case sc(L^R) where DFA for L is without dead state	with dead state	Upper bound $2^{n-2}+n-1$	Lower bound 2^{n-2}	Witness
1	1	1	1/2	1/2	
2	2	2	2	1	$L_2 = a^*b(a+b)^*$
3	4	4	4	2	$L_3 = b^* + b^*aL_2$
4	7	7	7	4	$L_4 = b^*aL_3$
5	12	12	12	8	$L_5 = b^*a(aL_3 + bL_2)$
6	21	21	21	16	$L_6 = b^*a(b^*a + L_5)$
7	**34**	34	**38**	32	$b^*ab^*a(a+b)(\varepsilon + aL_3 + bL_2)$
8	55	**64**	71	**64**	

Table 1. The tight bounds for the reverse of binary $\mathcal{R}$-trivial regular languages. The witness languages are $\mathcal{J}$-trivial.

$n \leq 7$, but not if $n = 8$. By Lemma 1, this means that for $n = 8$, the worst-case minimal partially ordered DFA has a dead state and so does its complement.

It is worth mentioning that the witness languages are even $\mathcal{J}$-trivial, hence these tight upper bounds also apply to binary $\mathcal{J}$-trivial regular languages discussed below.

We now prove that if $n \geq 8$, then the upper bound is 2^{n-2} in the case of binary alphabets.

Lemma 4. *Let $n \geq 8$ and let L be a binary $\mathcal{R}$-trivial regular language with* sc$(L) = n$. *Then* sc$(L^R) \leq 2^{n-2}$, *and the bound is tight.*

Proof. Consider a minimal partially ordered n-state DFA M over the alphabet $\{a, b\}$. If there are two maximal states, then one of them is accepting and the other is the dead state. Then the accepting state appears in all reachable subsets of the subset automaton of the NFA M^R, while the dead state appears in no reachable subset. Hence the upper bound is 2^{n-2}. If M has only one maximal state, and the state is the dead state, we take the complement that has the same state complexity by Lemma 1 and does not have the dead state.

Thus, it remains to investigate the upper bound for minimal partially ordered DFAs with no dead state. Assume that n is the maximal state. Note that if a minimal binary partially ordered DFA has at least four states, there is a path of

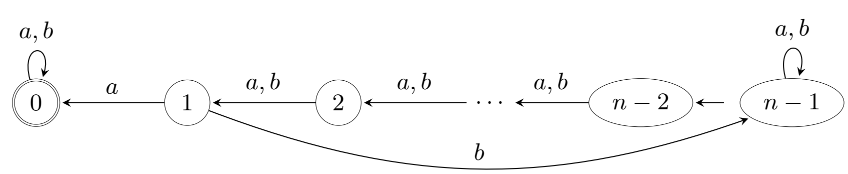

Fig. 2. A binary $\mathcal{R}$-trivial regular language meeting the bound 2^{n-2} for the reverse.

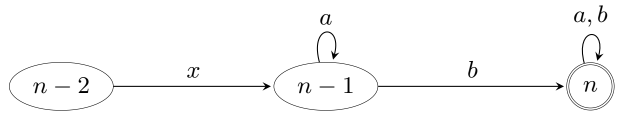

Fig. 3. Path of length two, where $x \in \{a, b\}$.

length two in the automaton. Consider three last states of such a longest path, say $n-2 \to n-1 \to n$. In particular, there is no longer path from $n-2$ to n. Note also that $n-1$ is not accepting, otherwise it is equivalent to n. As above in Theorem 2, we can show that to reach the upper bound, the situation must be as depicted in Fig. 3. Indeed, if both a, b go from $n-2$ to $n-1$, then there is at most one subset containing n and $n-1$ and not $n-2$, namely $F \cdot b$. Now, compute the number of reachable subsets containing n and $n-1$, but not containing $n-2$. Recall that $F \cdot a^k$, $k \geq 0$, reaches at most $n-1$ different subsets $X_1, \ldots, X_{n-1}$.

If $x = a$, then we have the following possibilities: (i) b goes to n; (ii) b goes to a state p from which one transition goes to n and one is a self-loop; (iii) b is a self-loop in $n-2$.

In the first case, there is no such subset because $n-1$ is introduced by b, which also introduces $n-2$.

In the second case, if p goes to n under a, then b is a self-loop in p, and there are at most $(n-2)$ subsets, $F \cdot b \cdot b^k$ with $k \geq 0$, containing n and $n-1$ and not containing $n-2$. If p goes to n under b, then a is a self-loop in p, and there are at most $(n-1)$ subsets, $F \cdot a^k \cdot b$ with $k \geq 0$, containing n and $n-1$ and not containing $n-2$.

In the third case, all the reachable subsets are $F \cdot a^k \cdot b \cdot b^\ell$ with $k, \ell \geq 0$. As there are at most $n-2$ subsets $F \cdot b \cdot b^\ell$, $Y_1, \ldots, Y_{n-2}$, the number of these subsets corresponds to the number of pairs $\{(X_i, Y_j) \mid i \geq j\}$, which is $((n-1)(n-2) + (n-2))/2 = n(n-2)/2$.

If $x = b$, then we have the following possibilities: (i) a goes to n; (ii) a goes to another state p from which one transition goes to n and one is a self-loop; (iii) a is a self-loop in $n-2$.

In the first case, there is only one subset containing n and $n-1$ and not containing $n-2$, namely $F \cdot b$.

In the second case, if p goes to n under a, then b is a self-loop in p and there are at most three subsets, $F \cdot b$, $F \cdot ba$ and $F \cdot ab$, containing n and $n-1$ and not containing $n-2$. If p goes to n under b, then a is a self-loop in p and there are at most $(n-1)$ subsets, $F \cdot a^k \cdot b$ with $k \geq 0$, containing n and $n-1$ and not containing $n-2$.

In the third case, all these subsets are reachable only by strings containing one b, that is, the subsets are reachable by $F \cdot a^k \cdot b \cdot a^\ell$. Their number is at most the number of pairs $\{(X_i, Z_j) \mid i \geq j\}$, where $F \cdot b \cdot a^\ell$ reaches subsets $Z_1, \ldots, Z_{n-2}$, which is $n(n-2)/2$.

Summarized, if $n \geq 4$, then the subset automaton of the NFA M^R has at most $2^{n-3} + \min(\frac{n(n-2)}{2}, 2^{n-3}) + (n-1)$ reachable states. Note that we get at

least 2^{n-2} subsets only in the case of $4 \leq n \leq 7$. Hence if $n \geq 8$, then the upper bound is reachable by automata with two maximal states, which means that the upper bound is 2^{n-2} if $n \geq 8$. ☐

Denote by $f_k(n)$ the state complexity function of the reverse on binary $\mathcal{R}$-trivial regular languages over a k-letter alphabet defined by

$$f_k(n) = \max\{\mathrm{sc}(L^R) \mid L \subseteq \Sigma^*, |\Sigma| = k, L \text{ is } \mathcal{R}\text{-trivial regular, and } \mathrm{sc}(L) = n\}.$$

Using this notation, we can summarize our results in the following theorem.

Theorem 2. *Let $n \geq 1$ and let $f_k(n)$ be the state complexity of the reverse on $\mathcal{R}$-trivial regular languages over a k-letter alphabet. Then*

$$f_1(n) = n,$$
$$f_2(n) = \begin{cases} 1, & \text{if } n = 1, \\ 2^{n-2} + n - 1, & \text{if } 2 \leq n \leq 6, \\ 34, & \text{if } n = 7, \\ 2^{n-2}, & \text{otherwise,} \end{cases}$$
$$f_3(n) = f_k(n) = 2^{n-1} \text{ for every } k \text{ with } k \geq 3.$$

Proof. Since the reverse of every unary language is the same language, we have $f_1(n) = n$. The upper bounds on f_2 are given by Lemmas 2 and 4 and by our calculations in the case of $n = 7$. The lower bounds in the case of $1 \leq n \leq 7$ also follow from the calculations, while the case of $n \geq 8$ is covered by Lemma 3. The result for f_3 is from Corollary 1. Since adding new symbols to the ternary witness automata does not change the proofs of reachability and distinguishability in the ternary case, the upper bound is tight for every k with $k \geq 3$. ☐

4 $\mathcal{J}$-trivial regular languages

As any $\mathcal{J}$-trivial regular language is $\mathcal{R}$-trivial, the upper bound for $\mathcal{R}$-trivial regular languages applies. To prove the results of this section, we first define *Simon's condition* on $\mathcal{R}$-trivial regular languages to be $\mathcal{J}$-trivial.

Any DFA $M = (Q, \Sigma, \delta, q_0, F)$ can be turned into a directed graph $G(M)$ with the set of vertices Q, where a pair $(p, q) \in Q \times Q$ is an edge in $G(M)$ if there is a transition from p to q in M. For $\Gamma \subseteq \Sigma$, we define the directed graph $G(M, \Gamma)$ with the set of vertices Q by considering only those transitions that correspond to letters in Γ. For two states p and q of M, we write $p \prec q$ if $p \preceq q$ and $p \neq q$. A state p is *maximal* if there is no state q such that $p \prec q$.

For a directed graph $G = (V, E)$ and $p \in V$, the set

$$C(p) = \{q \in V \mid q = p \text{ or there is a directed path from } p \text{ to } q\}$$

is called the component of p.

Definition 1 (Simon's condition). *A DFA M with an input alphabet Σ satisfies Simon's condition if, for every subset Γ of Σ, each component of $G(M, \Gamma)$ has a unique maximal state.*

Simon [17] has shown the following result.

Fact 3 *An $\mathcal{R}$-trivial regular language is $\mathcal{J}$-trivial if and only if its minimal partially ordered DFA satisfies Simon's condition.*

It is worth mentioning that it is more efficient to use Trahtman's condition to decide whether an $\mathcal{R}$-trivial regular language is $\mathcal{J}$-trivial. The improvement Trahtman has shown is the following: for a state p, let $\Sigma(p)$ denote the set of symbols under which there is a self-loop in p. Then an $\mathcal{R}$-trivial regular language is $\mathcal{J}$-trivial if and only if its minimal partially ordered DFA satisfies that, for every state p, the component of $G(M, \Sigma(p))$ containing p has a unique maximal state, see [19] for more details.

Concerning the lower bound state complexity, note that it was shown in [6] that there are finite languages over a binary alphabet whose reverse have a blow-up of $3 \cdot 2^{\frac{n}{2}-1} - 1$, for even n, and of $2^{\frac{n+1}{2}} - 1$, for odd n. As every finite language is $\mathcal{J}$-trivial, we obtain at least these lower upper bounds for binary $\mathcal{J}$-trivial regular languages.

Using Simon's result we immediately obtain the following lemma.

Lemma 5. *Let $\Gamma \subseteq \Sigma$. If a partially ordered DFA M over Σ satisfies Simon's condition, then the DFA M' (not necessarily connected) obtained from M by removing transitions under symbols from Γ also satisfies Simon's condition.*

Proof. Let $\Sigma' = \Sigma \setminus \Gamma$. By Fact 3, each component of $G(M, \Sigma')$ has a unique maximal state and remains partially ordered. □

We now prove the main result of this section.

Theorem 4. *At least $n-1$ symbols are necessary for a $\mathcal{J}$-trivial regular language of the state complexity n to reach the bound 2^{n-1} on the number of states in the reverse.*

Proof. We prove by induction on $|Q|$ that for every partially ordered DFA $M = (Q, \Sigma, \delta, q_0, F)$, where $|\Sigma| < |Q| - 1$, satisfying Simon's condition, the number of sets reachable in the subset automaton of the NFA M^R is strictly less than $2^{|Q|-1}$. The basis, $|Q| = 3$, holds since then the automaton is over a unary alphabet, which means that $L(M^R) = L(M)$, hence $\mathrm{sc}(L(M^R)) = \mathrm{sc}(L(M)) \leq 3 < 2^2$.

Assume that, for some $k \geq 3$, the claim holds for every partially ordered DFA satisfying Simon's condition with at most k states. Let M be a partially ordered DFA satisfying Simon's condition with $n = k + 1$ states, and assume that $2^{|Q|-1}$ subsets are reachable in the subset automaton of the NFA M^R. Without loss of generality, we may assume that the maximal state of M is denoted by n, and that it is accepting.

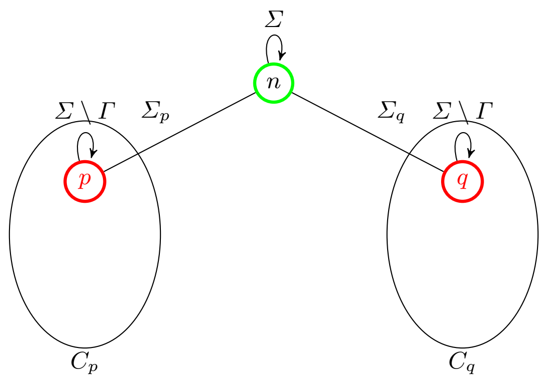

Fig. 4. The partially ordered DFA M'; $\Gamma = \Sigma_p \cup \Sigma_q$.

Let R denote the set all states different from n with a transition to n, that is, $R = \{q \in Q \setminus \{n\} \mid \delta(q,a) = n, a \in \Sigma\}$. Let Γ denote the set of symbols connecting the states of R with the state n, that is, $\Gamma = \{a \in \Sigma \mid \delta(R,a) \cap \{n\} \neq \emptyset\}$. Let M' be the k-state subautomaton of M obtained by removing the state n and all transitions labeled by symbols from Γ. In particular, the automaton M' satisfies Simon's condition by Lemma 5. Let $\max(R)$ be the set of all states of R that are maximal in M'. Note that every state of $\max(R)$ must be non-accepting, otherwise it appears in all reachable subsets of the subset automaton of M^R because of self-loops or transitions from n.

For a state p in $\max(R)$, let C_p denote the component of p in $G(M')$, and let $\Sigma_p = \{a \in \Sigma \mid \delta(p,a) = n\} \subseteq \Gamma$ denote the set of all labels under which p is connected to n, see Fig. 4 for illustration. Note that if C_p and C_q are connected, and $p \neq q$, then p and q are two maximal states of the component containing $C_p \cup C_q$, which is a contradiction because M' satisfies Simon's condition.

We now prove that for every state p in $\max(R)$, there exists a letter in Σ_p which does not appear in Σ_q for any other state q in $\max(R)$. For the sake of contradiction, assume that there is a state p in $\max(R)$ with $\Sigma_p \subseteq \bigcup_{q \in \max(R)}^{q \neq p} \Sigma_q$. Since all the subsets containing n and p and not containing any q different from p are reachable, and all states of $\max(R)$ are non-accepting, the state p is introduced to the subset from n by a transition under a letter from Σ_p which also introduces a state q into that subset. Since p and q are maximal, they cannot be later eliminated from the subset.

Thus, we have shown that for each state p in $\max(R)$, there exists a letter σ_p in Σ_p that does not appear in Σ_q for any state q in $\max(R)$ with $q \neq p$. Let $\Gamma' = \{\sigma_p \mid p \in \max(R)\}$ denote the set of these letters. Then $|\Gamma'| = |\max(R)|$.

Construct a new partially ordered DFA M'' from M' by joining all states of $\max(R)$ into one state. As they are all non-accepting, the subset automaton of the reverse of M'' has the same number of reachable states as the subset automaton of the reverse of M'. Moreover, as all sets that contain n and do not contain anything from $\max(R)$ are reachable in the subset automaton of M^R, the

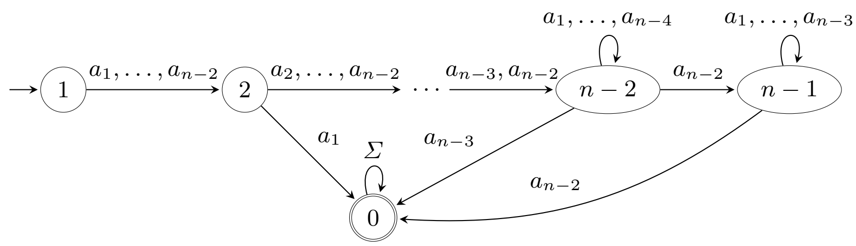

Fig. 5. The witness minimal partially ordered DFA M satisfying Simon's condition.

automaton M'' reaches the upper bound on the reverse. Thus, if $|\Sigma| < |Q| - 1$, then we have in M'' that $|\Sigma''| = |\Sigma| - |\max(R)| < |Q| - |\max(R)| - 1 = |Q''| - 1$. But, by the induction hypothesis, the automaton M'' cannot reach the upper bound on the reverse, which is a contradiction. $\square$

Using this result, we can prove the tight upper bound on the state complexity of the reverse for $\mathcal{J}$-trivial regular languages over an $(n-2)$-letter alphabet.

Theorem 5. *Let $n \geq 3$ and let L be a $\mathcal{J}$-trivial regular language over an $(n-2)$-letter alphabet with $\mathrm{sc}(L) = n$. Then $\mathrm{sc}(L^R) \leq 2^{n-1} - 1$, and the bound is tight.*

Proof. The upper bound follows from Theorem 4. To prove tightness, consider the $\mathcal{J}$-trivial regular language accepted by the minimal DFA $M = (\{0, 1, \ldots, n-1\}, \{a_1, \ldots, a_{n-2}\}, \delta, 1, \{0\})$ depicted in Fig. 5. The transitions under a symbol a_j in Σ are defined by

$$\delta(i, a_j) = \begin{cases} i+1, & \text{if } i \leq j \leq n-2, \\ 0, & \text{if } i = j-1, \\ i, & \text{otherwise.} \end{cases}$$

In the subset automaton of the NFA M^R, the initial state is state $\{0\}$ and, if $1 \leq k \leq n-2$, then every $(k+1)$-element set $\{0, i_1, i_2, \ldots, i_k\}$ with $2 \leq i_1 < i_2 < \cdots < i_k \leq n-1$ is reached from the k-element set $\{0, i_2, \ldots, i_k\}$ by a_{i_1}. This gives 2^{n-2} reachable states. Note that the subset $\{0, 1\}$ is not reachable, but all subsets of cardinality at least three containing 0 and 1 are reachable since every set $\{0, 1, i_1, \ldots, i_k\}$ is reached from the set $\{0, 2, i_2, \ldots, i_k\}$ by the symbol a_{i_1}. This completes the proof. $\square$

By Theorem 4, the upper bound for binary $\mathcal{J}$-trivial regular languages is at most $2^{n-3} + \min(\frac{n(n-2)}{2}, 2^{n-3}) + (n-1)$. Moreover, the witness languages presented in Table 1 are $\mathcal{J}$-trivial. Note that if $n \geq 8$, we need a dead state to reach the upper bound 2^{n-2}. Hence it is an open problem what is the tight upper bound for binary $\mathcal{J}$-trivial regular languages in the case of $n \geq 8$.

Corollary 2. *Let L be a binary $\mathcal{J}$-trivial regular language with $\mathrm{sc}(L) = n$, where $n \geq 4$, then $\mathrm{sc}(L^R) \leq 2^{n-3} + \min(\frac{n(n-2)}{2}, 2^{n-3}) + (n-1)$. A few tight upper bounds for small n are given in Table 1.*

5 Conclusions

We have characterized tight upper bounds on the state complexity of the reverse of $\mathcal{R}$-trivial regular languages depending not only on the state complexity of the language, but also on the size of the alphabet. This characterization also gives upper bounds for $\mathcal{J}$-trivial regular languages. However, we have shown that they are not reachable for $\mathcal{J}$-trivial regular languages of the state complexity n over an $(n-k)$-element alphabet, where $2 \leq k \leq n-3$. Namely, we have shown the tight upper bounds for $(n-1)$- and $(n-2)$-element $\mathcal{J}$-trivial regular languages, but (except for a few examples for binary $\mathcal{J}$-trivial regular languages) the problem what are the tight upper bounds for $\mathcal{J}$-trivial regular languages over an alphabet of a lower cardinality is open.

Acknowledgements. The authors gratefully acknowledge very useful suggestions and comments of anonymous referees on the previous version of this work.

References

1. Bojanczyk, M., Segoufin, L., Straubing, H.: Piecewise testable tree languages. Logical Methods in Computer Science 8(3) (2012)
2. Brzozowski, J.A.: Canonical regular expressions and minimal state graphs for definite events. In: Symposium on Mathematical Theory of Automata. MRI Symposia Series, vol. 12, pp. 529–561. Polytechnic Institute of Brooklyn, New York (1963)
3. Brzozowski, J.A., Fich, F.E.: Languages of $\mathcal{R}$-trivial monoids. Journal of Computer and System Sciences 20(1), 32–49 (1980)
4. Brzozowski, J.A.: Quotient complexity of regular languages. In: DCFS. EPTCS, vol. 3, pp. 17–28 (2009)
5. Brzozowski, J.A., Li, B.: Syntactic complexity of $\mathcal{R}$- and $\mathcal{J}$-trivial regular languages. CoRR ArXiv 1208.4650 (2012)
6. Câmpeanu, C., Culik II, K., Salomaa, K., Yu, S.: State complexity of basic operations on finite languages. In: WIA. LNCS, vol. 2214, pp. 60–70. Springer (1999)
7. Czerwiński, W., Martens, W., Masopust, T.: Efficient separability of regular languages by subsequences and suffixes. CoRR ArXiv 1303.0966 (2013)
8. Jahn, F., Kufleitner, M., Lauser, A.: Regular ideal languages and their boolean combinations. In: CIAA. LNCS, vol. 7381, pp. 205–216. Springer (2012)
9. Jirásková, G., Masopust, T.: Complexity in union-free regular languages. International Journal of Foundations of Computer Science 22(7), 1639–1653 (2011)
10. Jirásková, G., Masopust, T.: On the state and computational complexity of the reverse of acyclic minimal DFAs. In: CIAA. LNCS, vol. 7381, pp. 229–239. Springer (2012)
11. Jirásková, G., Klíma, O.: Descriptional complexity of biautomata. In: DCFS. LNCS, vol. 7386, pp. 196–208. Springer (2012)
12. Klíma, O., Polák, L.: On biautomata. In: NCMA. books@ocg.at, vol. 282, pp. 153–164. Austrian Computer Society (2011)
13. Lawson, M.: Finite Automata. Chapman and Hall/CRC (2003)
14. Leiss, E.: Succinct representation of regular languages by boolean automata. Theoretical Computer Science 13, 323–330 (1981)

15. Rogers, J., Heinz, J., Bailey, G., Edlefsen, M., Visscher, M., Wellcome, D., Wibel, S.: On languages piecewise testable in the strict sense. In: MOL. LNCS, vol. 6149, pp. 255–265. Springer (2009)
16. Simon, I.: Hierarchies of Events with Dot-Depth One. Ph.D. thesis, Dep. of Applied Analysis and Computer Science, University of Waterloo, Canada (1972)
17. Simon, I.: Piecewise testable events. In: GI Conference on Automata Theory and Formal Languages. pp. 214–222. Springer (1975)
18. Stern, J.: Complexity of some problems from the theory of automata. Information and Control 66(3), 163–176 (1985)
19. Trahtman, A.N.: Piecewise and local threshold testability of DFA. In: FCT. LNCS, vol. 2138, pp. 347–358. Springer (2001)
20. Yu, S., Zhuang, Q., Salomaa, K.: The state complexities of some basic operations on regular languages. Theoretical Computer Science 125(2), 315–328 (1994)

Appendix

Fact 1. All states of the subset automaton corresponding to the reverse of a minimal DFA M are pairwise distinguishable.

Proof. Let q be a state of the NFA M^R. Since q is reachable in M, there exists a string w_q which is accepted by M^R from the state q. Moreover, the string w_q cannot be accepted from any other state of M^R because M is deterministic. Next, two distinct subsets of the subset automaton differ in a state q of M^R, and the string w_q distinguishes these two subsets. □

Figures from Corollary 1

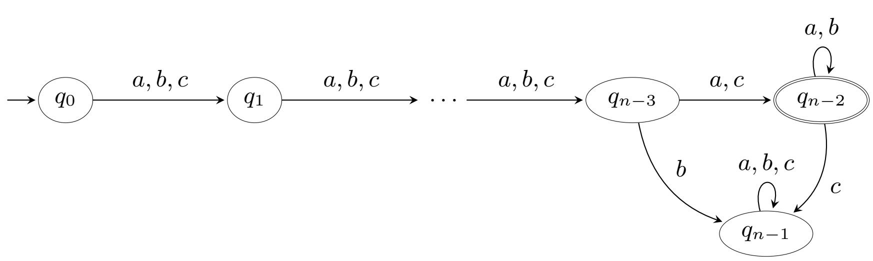

Fig. 6. Figure of [10, Lemma 3, p. 232].

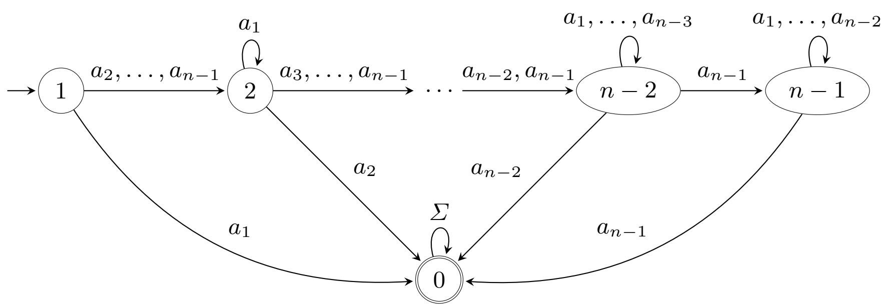

Fig. 7. Figure of [5, v2, Theorem 5, p. 15].